

## Artikel erschienen in:

Ottmar Ette, Eberhard Knobloch (Hrsg.)

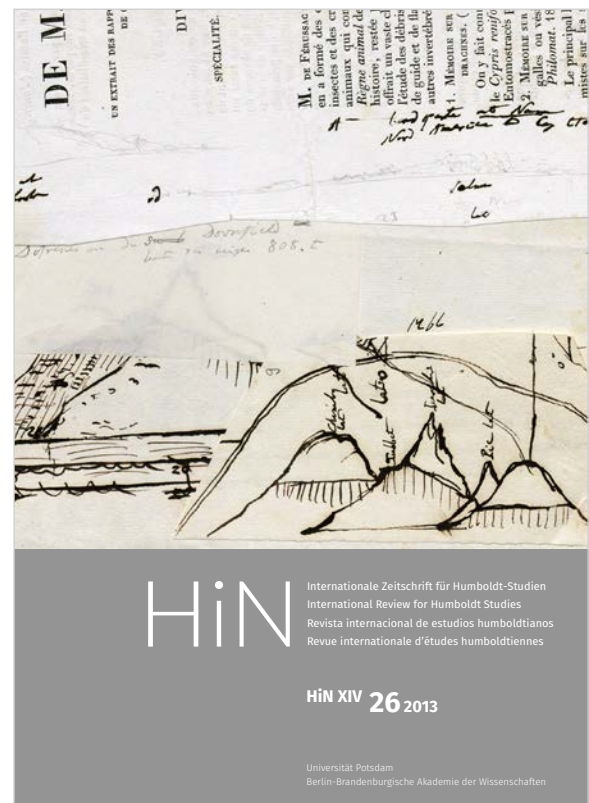
### HiN : Alexander von Humboldt im Netz, XIV (2013) 26

2013 – 70 p.

ISSN (print) 2568-3543

ISSN (online) 1617-5239

URN urn:nbn:de:kobv:517-opus-66611



#### Empfohlene Zitation:

Detlev Doherr: Interconnectedness und digitale Texte, In: Ette, Ottmar; Knobloch, Eberhard (Hrsg.). HiN : Alexander von Humboldt im Netz, XIV (2013) 26, Potsdam, Universitätsverlag Potsdam, 2013, S. 13–19.  
DOI <https://doi.org/10.18443/174>

Soweit nicht anders gekennzeichnet ist dieses Werk unter einem Creative Commons Lizenzvertrag lizenziert: Namensnennung 4.0. Dies gilt nicht für zitierte Inhalte anderer Autoren:  
<https://creativecommons.org/licenses/by-nc/4.0/deed.de>



# Detlev Doherr

## Interconnectedness und digitale Texte

---

### *Zusammenfassung*

Die multimedialen Informationsdienste im Internet werden immer umfangreicher und umfassender, wobei auch die nur in gedruckter Form vorliegenden Dokumente von den Bibliotheken digitalisiert und ins Netz gestellt werden. Über Online-Dokumentenverwaltungen oder Suchmaschinen können diese Dokumente gefunden und dann in gängigen Formaten wie z.B. PDF bereitgestellt werden.

Dieser Artikel beleuchtet die Funktionsweise der Humboldt Digital Library, die seit mehr als zehn Jahren Dokumente von Alexander von Humboldt in englischer Übersetzung im Web als HDL (Humboldt Digital Library) kostenfrei zur Verfügung stellt. Anders als eine digitale Bibliothek werden dabei allerdings nicht nur digitalisierte Dokumente als Scan oder PDF bereitgestellt, sondern der Text als solcher und in vernetzter Form verfügbar gemacht.

Das System gleicht damit eher einem Informationssystem als einer digitalen Bibliothek, was sich auch in den verfügbaren Funktionen zur Auffindung von Texten in unterschiedlichen Versionen und Übersetzungen, Vergleichen von Absätzen verschiedener Dokumente oder der Darstellung von Bildern in ihrem Kontext widerspiegelt.

Die Entwicklung von dynamischen Hyperlinks auf der Basis der einzelnen Textabsätze der Humboldt'schen Werke in Form von Media Assets ermöglicht eine Nutzung der Programmierschnittstelle von Google Maps zur geographischen wie auch textinhaltlichen Navigation.

Über den Service einer digitalen Bibliothek hinausgehend, bietet die HDL den Prototypen eines mehrdimensionalen Informationssystems, das mit dynamischen Strukturen arbeitet und umfangreiche thematische Auswertungen und Vergleiche ermöglicht.

### *Summary*

The multimedia information services on Internet are becoming more and more comprehensive, even the printed documents are digitized and republished as digital Web documents by the libraries. Those digital files can be found by search engines or management tools and provided as files in usual formats as PDF.

This article focuses on the functionality of the Humboldt Digital Library (HDL), which is available for more than a decade and which provides digitized documents of Alexander von Humboldt in the English translations. Unlike a digital library, the HDL doesn't provide only digitized documents as PDF or image formats with book scans, but the text itself is made available as such and embedded into an information network. Therefore the system is representing an information system rather than a digital library, which is also reflected by the implemented functions for the search of texts in different versions and translations, comparisons of paragraphs of different documents or presentation of images in their contexts. The development of dynamic hyperlinks, based on the individual paragraphs in the Humboldt works in the form of media assets, enables the use of the Application Programming Interface (API) of Google Maps for geographical as well as text content navigation. Beyond the services of a digital library the HDL provides a prototype of a multi-dimensional information system that works with dynamic structures and enables extensive thematic research and comparisons.

# Detlev Doherr

## Interconnectedness und digitale Texte

---

### *Resumen*

Los servicios de información multimedia en Internet se están convirtiendo cada vez más comprensivos, incluso los documentos impresos son digitalizados y republicados, por las bibliotecas, como documentos Web digitales. Estos archivos digitales pueden ser encontrados por máquinas de búsqueda o herramientas administrativas, y proveerlos como archivos en formatos comunes como PDF.

Este artículo se enfoca en la funcionalidad de la Biblioteca Digital Humboldt (HDL=Humboldt Digital Library), la cual está disponible por más de una década y provee documentos digitalizados de Alexander von Humboldt traducidos al inglés. A diferencia de una biblioteca digital, la HDL no provee solamente documentos digitales en PDF o en formatos de imagen de libros escaneados, sino el texto en sí mismo se pone a disposición como tal y embebido en una red de información. Por lo tanto, el sistema representa un sistema de información más que una biblioteca digital, lo cual se muestra también por las funciones implementadas para el descubrimiento de textos en diferentes versiones y traducciones, comparaciones de párrafos de diferentes documentos o la presentación de imágenes en su contexto.

El desarrollo de hyperlinks dinámicos basado en párrafos individuales de los trabajos de Humboldt, trabaja en la forma de media assets, permite el uso de Interface de Aplicación Programada (API=Application Programming Interface) de Google Maps para navegación tanto geográfica así como de contenido de texto. Más allá de servicios de biblioteca digital, la HDL proporciona un prototipo de un sistema de información multi-dimensional que funciona con estructuras dinámicas y habilita vastas evaluaciones y comparaciones temáticas.

## Entwicklungen der HDL

Die Entwicklung der HDL wurde im Jahr 2000 aufgenommen und fand im Rahmen von Transcoop- Projekten mit Drittmitteln der Alexander-von-Humboldt-Stiftung statt. In dieser Phase wurden die Anforderungen an eine Web-Präsentation der Werke von Humboldt erörtert und erste Internetauftritte organisiert (Abb. 1). In der folgenden Phase wurden die aus dem Transcoop-Projekt gewonnenen Erkenntnisse in weiterführenden Konzepten verdichtet und verschiedene Versuche mit den damals verfügbaren Digital Library-Plattformen (DSpace, EPrints) unternommen. Ergänzend zu den verfügbaren Funktionen wurden thematische Strukturen, Parallelisierungen von Versionen und Übersetzungen sowie Informationsvernetzungen über die Dokumentgrenzen hinweg geplant. Um diese ergänzenden Funktionen anbieten zu können, wurde in dieser Zeit mit der Systementwicklung der heutigen HDL begonnen.<sup>1</sup>

In der derzeitigen Entwicklungsphase ist der Prototyp der HDL seit längerer Zeit im Web verfügbar und hat etwa 500000 Hits pro Jahr. Die Web 2.0-Strukturen wurden für registrierte User freigeschaltet, so dass Book-

marking, Personal Notes und Blogs möglich sind. Die übergreifenden Analysefunktionen wie ein komplettes Wortverzeichnis der Humboldt-Literatur und eine daraus abgeleitete Autocomplete-Funktion bei der Eingabe von Schlagworten sind ebenso realisiert wie die Parallelisierung der verfügbaren Übersetzungen zu den englischen Texten. Die semantischen Strukturen und maschinenverwertbare Informationen werden erkennbar über die Relevanzbewertung der HDL für verschiedene Textabsätze, die in Abhängigkeit vom Verhalten aller registrierten User gesteuert wird.

Das System der HDL ist für jeden Internetnutzer kostenfrei nutzbar, eine Anmeldung am System erweitert die Bearbeitungsmöglichkeiten durch die selbstverwaltete Eingabe von Userprofilen mit speziellen Themenschwerpunkten für Suchvorgänge, Speicherung von Suchpfaden und verwendeten Textstellen sowie der Systemkalkulation von Textbausteinen bezüglich der Relevanz für Suchbegriffe und thematische Suchen.

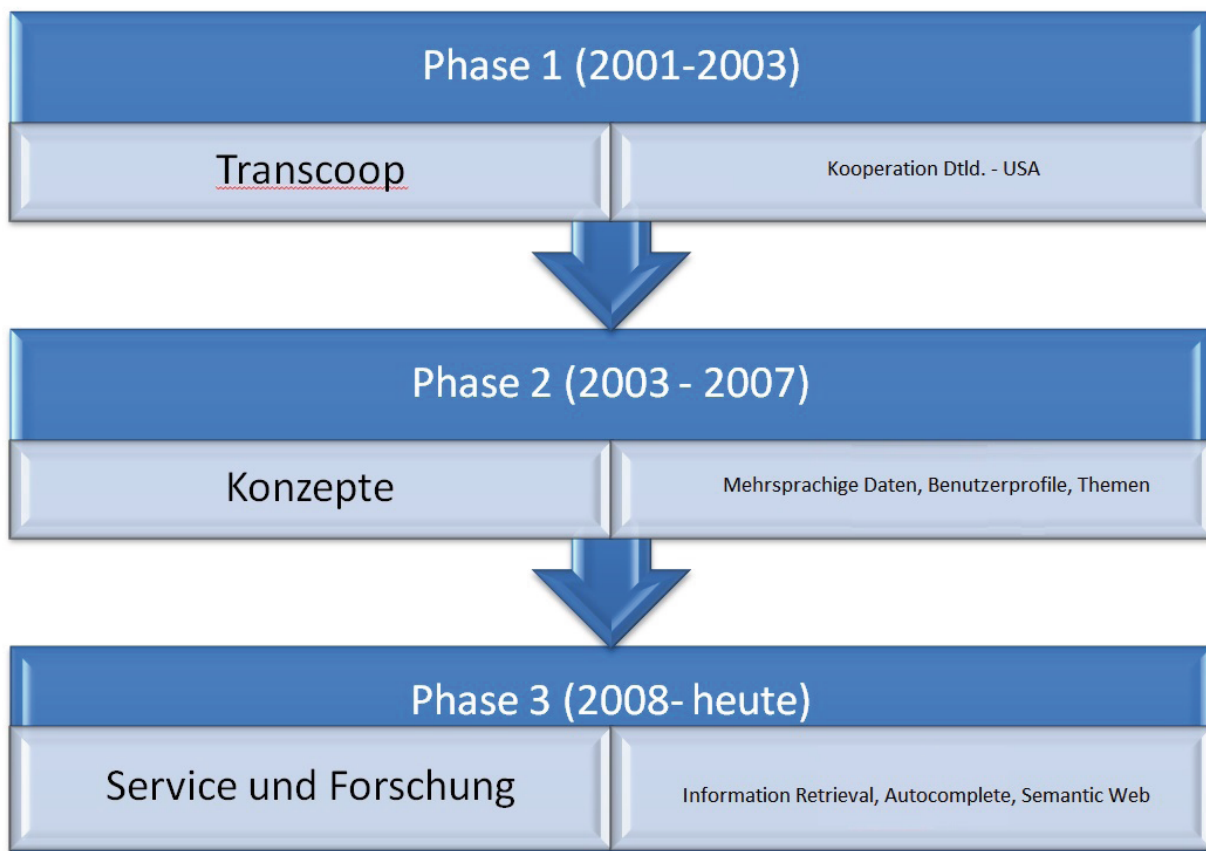


Abb.1: Entwicklung der HDL in den einzelnen Phasen seit dem Jahre 2000

<sup>1</sup> Doherr, D. (2005): The Humboldt Digital Library: Exploring Innovative Structures. *Humboldt im Netz (HiN)*, Sondernummer VI,10, ISSN 1617-5239.

## Aufbau der HDL und interne Struktur

Auf der Grundlage der Humboldt'schen Dokumente, die als Scans und über automatische Texterkennung gewonnene Texte im System hinterlegt sind, wurde eine Internet- Plattform geschaffen, die mehrschichtig aufgebaut ist.<sup>2</sup> Wie in Abbildung 2 dargestellt, geschieht der Zugriff auf das System über die Website, die die Webseiteninhalte mit einem Content-Management-System verwaltet und den Zugriff auf die Systemfunktionen bietet. Die Inhalte aus der Datenbank werden über eine Schnittstelle (Application Interface) bereitgestellt, mit der entweder eine Suchmaschine (Information Retrieval) oder eine Dokumentenanzeige (Presentation Manager) genutzt werden kann. Auf dieser Ebene ist auch die Google-Funktionalität über die Programmierschnittstelle Google API implementiert. Die Datenbank enthält neben den absatzorientierten Texten auch die Illustrationen, Karten und Scans der Originadokumente, die je nach Anforderung präsentiert werden. Zusätzlich dazu dienen zahlreiche Systemtabellen der relationalen Datenbank dazu, die Vernetzung der Textinformationen und deren thematische und funktionelle Zuordnungen zu realisieren. Zur Analyse von Suchpfaden, Relevanzberechnungen und userbezogenen Diensten dient eine Prozessdatenbank, die die entsprechenden Vorgänge in der HDL registriert und systemtechnisch auswertet. Diese prozessorientierte Komponente ist im Augenblick in der Entwicklungsphase und wird in der aktuellen Version der HDL getestet.

Die HDL ist im Rahmen der Open Archive Initiative (OAI) als Content Provider registriert, womit das System als Informationsquelle zu Humboldts Werken weltweit genutzt werden kann. Obwohl die HDL eine von XML unabhängige interne Struktur von Daten verwaltet, können die Dokumente gemäß der Text Encoding Initiative (TEI)<sup>3</sup> mit den generischen Richtlinien zur Darstellung von Texten auch auf XML-Basis dargestellt werden. Das Media Asset Management erlaubt es, die ver-

schiedenen Textbausteine auch in mehreren Versionen oder Übersetzungen zu verwalten, wobei die Parallelisierung der Absätze bei der Datenübernahme aus digitalisierten Quellen systemgestützt erfolgt. Diese auf den Media Assets aufbauenden Informationsvernetzungen können allerdings mit der statischen Form der Textspeicherung im XML Format nicht ausreichend repräsentiert werden.

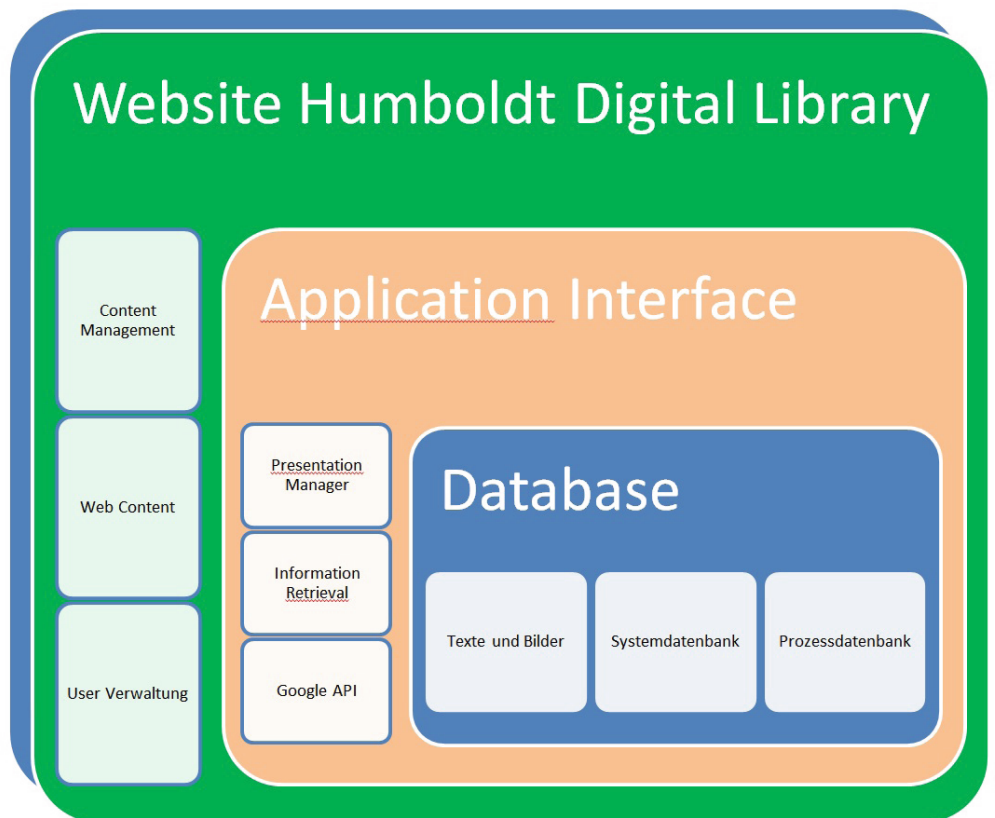


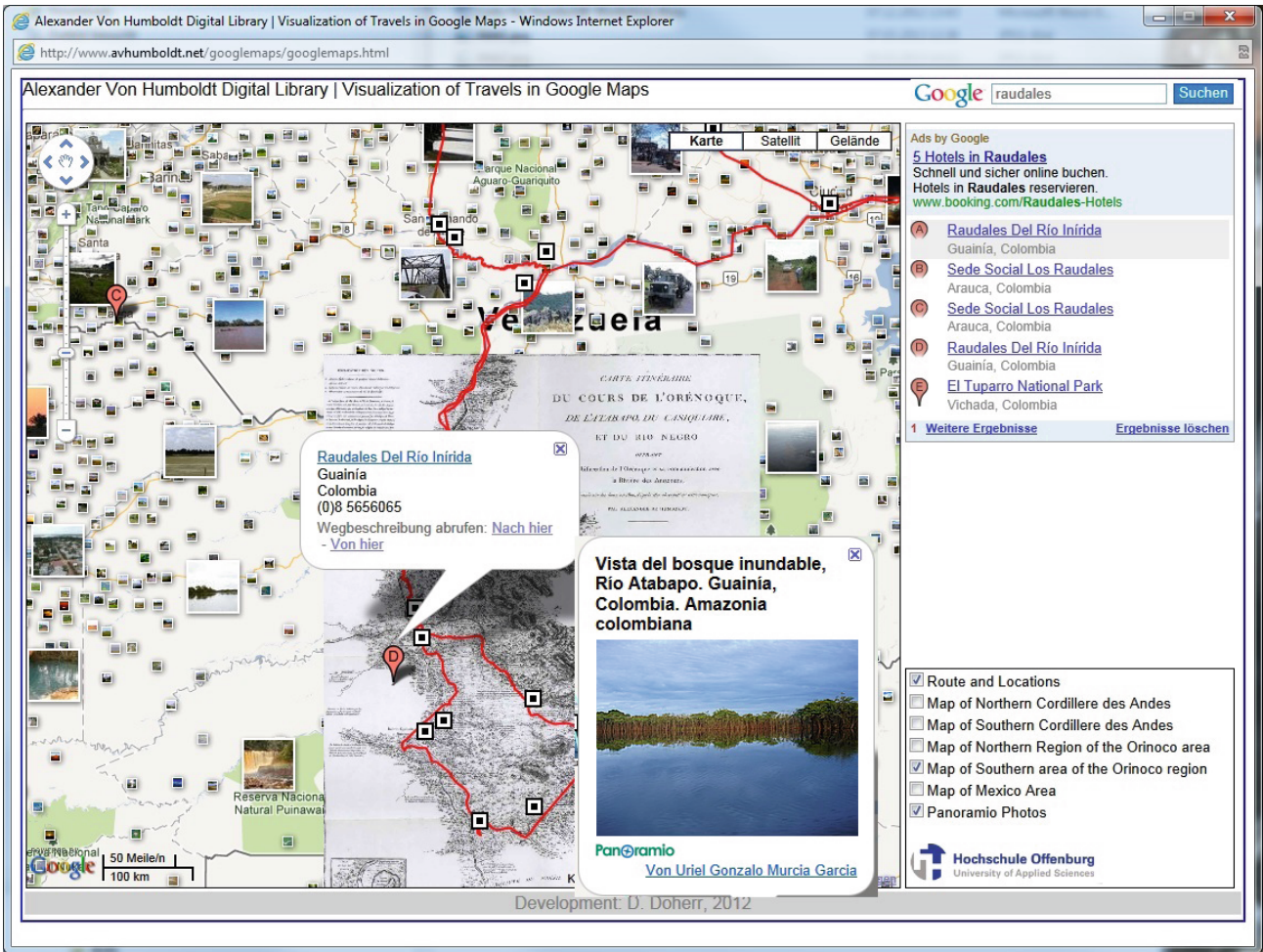
Abb. 2: Interne Struktur der HDL und Ebenenkonzept mit Zugriff über die Website und Anwendungsschnittstelle (Application Interface) zu den strukturierten Datenbeständen von Inhalten, System- und Prozessdatenbank

## Google Maps und Google Earth-Implementierung

Weit über Reiseliteratur hinaus beschreibt Humboldt verschiedenste Naturerscheinungen und Beobachtungen an den unterschiedlichsten Orten der Erde. Um die Reisewege sichtbar zu machen, aber auch die Reise und die Schlussfolgerungen virtuell nachvollziehen zu können, muss eine Verknüpfung der Texte mit den geographisch orientierten Informationen vorgenommen werden. Dieses kann die HDL über die Verknüpfung von Google Earth und Google Maps mit der Textdatenbank liefern. Darüber hinaus kann das Kartenmaterial als Overlay Map auf der Grundlage der im Google verfügbaren geographischen Daten georeferenziert und die Beobachtungen Humboldts mit heutigen Verhältnissen

<sup>2</sup> Doherr, D. & Baron, F. (2011): Humboldt Digital Library and Interconnectedness. *Environmentalist*, 32, S. 271-277, DOI 10.1007/s 10669-011-9369-y, Springer-Verlag 2012

<sup>3</sup> TEI: <http://www.tei-c.org/index.xml>; Zugriff am 05.03.2013



verglichen werden.<sup>4</sup> Die Verbindung zu den Google-Diensten ist technisch realisiert über einen XML-Parser, der die Metadaten der HDL in eine virtuelle XML-Datei transformiert, analysiert und an die Google API übergibt.

Google Maps braucht keine lokale Installation auf dem Rechner eines HDL-Nutzers und kann deshalb plattformunabhängig in jedem Browser genutzt werden. Die auf JavaScript-Programmierung basierende Google Maps API (Application Programming Interface)<sup>5</sup> stellt eine Reihe von Funktionen zur Entwicklung individueller Google Maps Anwendungen zur Verfügung. Dazu gehört die Objektdefinition im sog. KML Format (Keyhole Markup Language), welches ein XML Schema zur Darstellung von geographischen Daten ist.<sup>6</sup> Dieses XML Schema wird in der HDL benutzt, um die Reiserouten, das Kartenmaterial als Overlay Maps und die

Abb. 3: Die Verknüpfung von Google Maps und den Humboldtschen Texten aus der HDL ermöglicht eine Wegverfolgung der Reiseroute von Humboldt, aber auch eine dynamische Verknüpfung der Texte mit den jeweiligen Orten (und häufig auch der Zeiten). Außerdem sind die aktuellen Ortsdaten aus den Google-Diensten (hier in roten Placemarks) im Vergleich mit den Humboldt-Daten direkt verfügbar.

Ortsbeschreibungen als Placemarks zu definieren. Aber auch die Einbeziehung externer Dienste wie z.B. Panoramio (Photo-Sharing-Community) als Erweiterung der visuellen Präsentation von Humboldts Reisezielen ist möglich und realisiert. Die Darstellung der Karte kann in vollem Umfang angepasst werden und bietet dem Benutzer verschiedene Kartentypen (Straßen/Satellit/Gelände).

Ein besonderes Merkmal der Einbindung in die HDL ist es, über die Ortsmarkierungen mit einem direkten Link zu den Humboldt-Dokumenten Einblick in die Arbeit Humboldts zu bekommen. Dazu dienen die dynamischen Hyperlinks, die anders als die im Internet üblichen Hyperlinks stets auf die gleiche Internetadresse verweisen, aber die Suchmaschine mit Hilfe von dynamisch übertragenen Metadaten nur die Texte präsentiert, die zu den jeweiligen Orten passen. Die Kriterien

4 Vgl. Doherr, D. (2005): The Humboldt Digital Library: Exploring Innovative Structures.

5 <https://developers.google.com/maps> und [../kml/documentation/kml-reference](https://developers.google.com/maps/documentation/kml/); Zugriff: 04.03.2013

6 <https://developers.google.com/kml/documentation/kmlreference>; Zugriff am 05.03.2013

für die Textsuche sind im einfachen Falle Schlüsselworte oder Identifikationsnummern für Textbausteine. Aber auch thematische oder geographische Suchkriterien können zum Auffinden der Texte als Filter genutzt werden.

Die dynamischen Hyperlinks werden in den Beschreibungen der jeweiligen Ortsmarkierungen unter den jeweils präsentierten Textbausteinen angeboten und führen bei Aktivierung über das Information Retrieval Modul der HDL zu einer oder mehreren damit verbundenen Textpassagen in den unterschiedlichen Dokumenten.

Zusätzlich gestaltet sich die Einbindung der Google-Dienste auf der HDL-Webseite als nützlicher Service, um Vergleiche zwischen den Humboldt'schen Beschreibungen und den aktuellen Situationen durchzuführen. Damit navigiert der Anwender virtuell auf den Reiserouten Humboldts und kann diese Reise in der Datenbank jeweils nachvollziehen (Abb. 4). Auch die Vergleiche zu aktuellen Ortsbeschreibungen und umfangreiches Bildmaterial (Panoramio) sind direkt im selben Browserfenster möglich.

## Nachhaltigkeit

Humboldt brachte seine Beobachtungen stets in einen Zusammenhang mit den natürlichen Entwicklungen mit zeitlicher Dimension und gegenseitigen Wechselwirkungen. Und als Geowissenschaftler, der u.a. in Göttingen studiert hatte, konnte Humboldt die Wechselwirkungen zwischen belebter und unbelebter Natur als dynamische Prozesse erkennen und beschreiben. Mit diesem Kontext der Beobachtungen und der historischen Dimension legte Humboldt u. a. den Grundstein für die sehr moderne Vorstellung von Nachhaltigkeit und nachhaltiger Entwicklung, die in unserer heutigen Gesellschaft eine bedeutende Rolle spielt. Im politischen Sinne fordert dieses Konzept, das menschliche Handeln auf mehr Generationengerechtigkeit und globale Chancengleichheit auszurichten, um den Ressourcenverbrauch einzugrenzen und die Umwelt nicht zu schädigen. Wie Lucht 2009 referierte, hat sich „Alexander von Humboldt... unter anderem gefragt, wie eine Wissenschaft der Erde aussehen müsste, welche die Wechselwirkungen zwischen der Erde, dem Leben und dem Menschen darstellt“.<sup>7</sup>

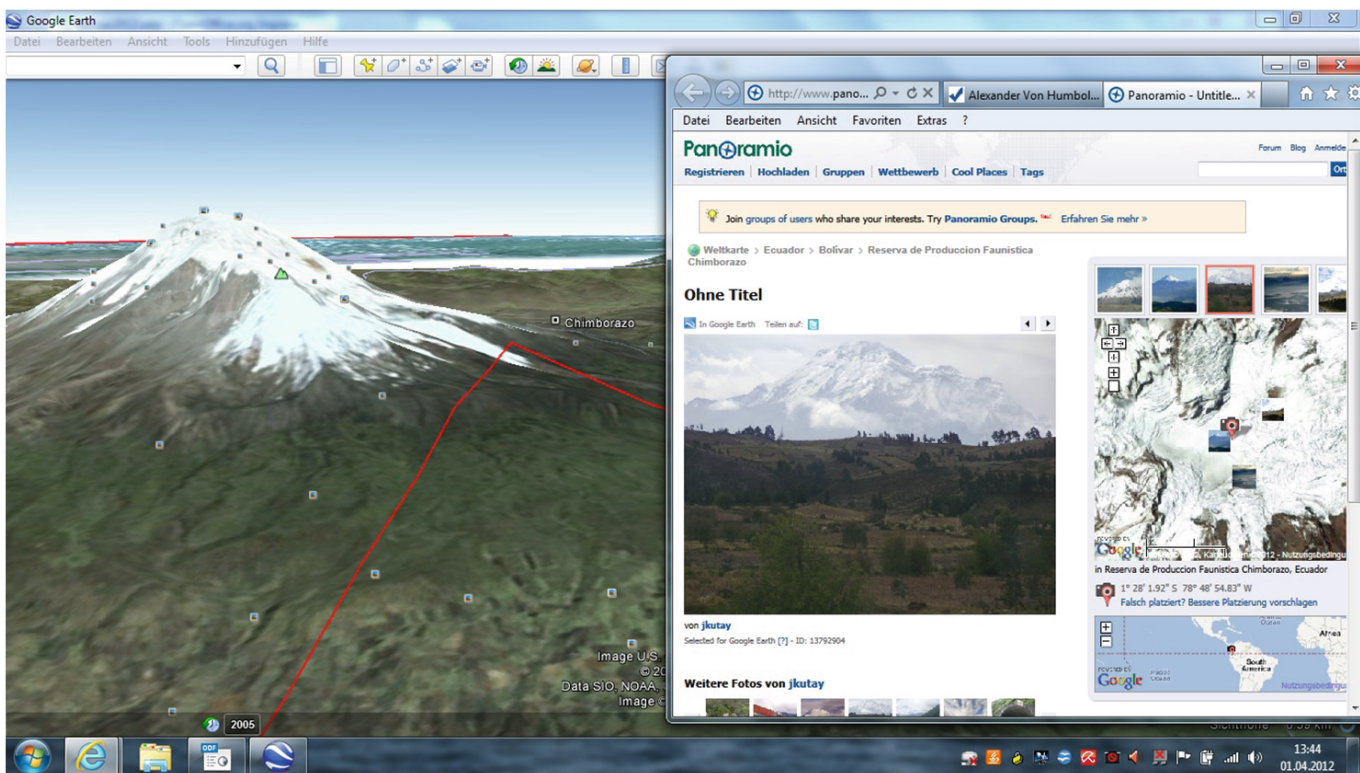


Abb. 4: Das Google Earth bietet eine 3D-Perspektive der Erdoberfläche, die in der HDL auch die Humboldt'schen Reiseroute und die Beobachtungspunkte enthält. Die Google-Dienste erlauben den direkten Durchgriff auf aktuelle Informationen wie z.B. die Bilder der Panoramio Community, die zu jedem der Reiseorte Humboldts verfügbar sind.

Für digitale Bibliotheken ergeben sich verschiedene Konsequenzen für die Datensicherung und Bereitstellung der Humboldt'schen Werke angesichts der Beschreibungen komplexer Prozesse und Zusammen-

7 Lucht, W. (2009): <http://www.hu-berlin.de/pr/medien/publikationen/humboldt/2008/200905/thema/krisen/>; Zugriff am 05.03.2013



hänge in thematischen und zeitlichen Dimensionen. So kann das Humboldt'sche Konzept der Wechselwirkungen im Zeitalter des Internets nur mit dynamisch vernetzten Daten adäquat abgebildet werden.<sup>8</sup> Das Internet und die moderne Informationstechnik bieten uns über die Methodik der Informationsverlinkung vielfältige Möglichkeiten, komplex vernetzte Daten zu analysieren, strukturieren und zu präsentieren. Jedoch kommt es immer stärker auf die Methodik der Informationsgewinnung und -bereitstellung an. Dieses ist immer seltener ein singulärer Text und/oder Bild, sondern vielmehr ein zusammenhängendes Kontinuum aus dynamisch aggregierten Informationselementen, die in einer virtuellen Umgebung innerhalb einer Website abgebildet werden. Die vielschichtigen Werke Alexander von Humboldts implizieren die Bereitstellung dynamischer Informationssysteme in einem semantischen Web in welchem Informationen vom Computersystem dynamisch aggregiert und je nach Suchanforderung bereitgestellt werden. Das Information Retrieval Modul der HDL erfüllt bereits einige der geforderten Funktionalitäten wie z.B. die Autocomplete-Funktion zur Unterstützung bei der Schlagwortsuche oder die dynamischen Hyperlinks, mit denen die Verlinkung von Google Maps und Google Earth implementiert wurde.

## Ausblick

Die HDL ist im derzeitigen Entwicklungsstand ein Prototyp für eine virtuelle Forschungsplattform mit dynamischen Strukturen. Die Qualität der verfügbaren Daten ist gemäß der zur Verfügung stehenden Ressourcen verbesserungswürdig, da das Projekt in erster Linie ursprünglich nur auf einen eng begrenzten Umfang aus englischen Texten ausgelegt war. Über die Integration von autorisierten Dokumenten und Bibliotheksbeständen können die Funktionalitäten der HDL aber auf alle Humboldt-Werke in allen verfügbaren Übersetzungen ausgeweitet werden, da die Systemressourcen dieses zulassen. Eine entsprechende Datenintegration kann über autorisierte XML-Dokumente erreicht werden.

Weiterhin wird zurzeit daran gearbeitet, ein Portal für alle online verfügbaren Humboldt-Werke zu schaffen. Diese Portalfunktion wird in einem eigenständigen Projekt in Kürze fertig gestellt sein und im Internet freigeschaltet werden. Es wird z.B. alle im Google Books verfügbaren Werke im Portal erschließen und online erreichbare Bibliotheksbestände abbilden. Allerdings kann eine systemtechnische Vernetzung der Textbausteine wie in der HDL und eine Zuordnung zu thematischen Kriterien bei den externen Daten bisher nur

ansatzweise realisiert werden. Daher besteht nach wie vor das Bestreben, verfügbare Texte auch direkt in die HDL einzubinden, um den Vernetzungsgrad innerhalb der Humboldt'schen Werke zu erhöhen. Von einer automatisierten Informationsverarbeitung innerhalb eines semantischen Web sind wir aber noch ein gutes Stück entfernt.

\* \* \*

## Zitierweise

Doherr, Detlev (2013): Interconnectedness und digitale Texte. In: *HiN - Humboldt im Netz. Internationale Zeitschrift für Humboldt-Studien* (Potsdam - Berlin) XIV, 26, S. 12-18. Online verfügbar unter <<http://www.uni-potsdam.de/u/romanistik/humboldt/hin/hin26/doherr.htm>>

Permanent URL unter <[http://opus.kobv.de/ubp/abfrage\\_collections.php?coll\\_id=594&la=de](http://opus.kobv.de/ubp/abfrage_collections.php?coll_id=594&la=de)>

<sup>8</sup> Doherr, D. & Baron, F. (2011): Humboldt Digital Library and Interconnectedness.